



ELSEVIER

Contents lists available at ScienceDirect

Language Sciences

journal homepage: www.elsevier.com/locate/langsci

The similar rates of occurrence of consonants across the world's languages: A quantitative analysis of phonetically transcribed word lists

Caleb Everett

University of Miami, Anthropology Department, P.O. Box 248106, Coral Gables, FL, 33124-2005, USA

ARTICLE INFO

Article history:

Received 30 November 2017

Received in revised form 3 May 2018

Accepted 8 July 2018

Available online xxx

Keywords:

Phonetics

Phonology

Zipf

Typology

Ease of articulation

ABSTRACT

Via an analysis of nearly 7000 phonetically transcribed word lists representing every major linguistic family, this study examines the frequency with which languages use particular consonants. The results suggest that there is an underlying global similarity in the frequency of occurrence of these consonants. Some consonants are uncommon across the word lists, and do not occur frequently even in those languages in which they are used. In contrast, a few other consonants represent a large portion of sounds in the word lists, across regions and linguistic families. The analysis quantifies the pervasiveness of such sounds that are known to be, impressionistically, common in speech. A new method captures the overall similarity of consonants' rates of occurrence across languages and families. This method offers one means of evaluating the extent to which individual languages or families deviate from, or adhere to, the typical patterns of consonant usage. It is suggested that the crosslinguistic similarity of consonant usage, in terms of rates of occurrence in word lists, likely owes itself to previously documented factors like the relative ease of articulation of some sounds. This new evidence suggests that the role of ease of articulation in shaping speech may be more influential than generally assumed, though other explanatory factors are also likely at work and these patterns require further exploration with more robust intralinguistic phonetic corpora.

© 2018 Elsevier Ltd. All rights reserved.

1. Introduction

The world's linguistic diversity includes an impressive array of sounds used to make speech happen (Maddieson, 1984; Evans and Levinson, 2009). Languages have from 11 to 141 phonemes, judging from surveys of the extant phoneme databases, and the kinds of sounds used as phonemes are well known to vary substantially (Gordon, 2016). Additionally, languages vary with respect to permissible syllable structures and sundry other articulatory phenomena. The sounds used in speech are, simply, remarkably diverse. In principle, languages could also vary dramatically with respect to the frequency with which they use the individual sounds at their disposal. Yet it is unclear just how diverse languages are in terms of how much they rely on particular sounds. Most work on the typology of sound systems focuses on the composition of phoneme inventories, not on the relative frequency in speech of individual phonemes and their variants. These inventories have played a critical role in offering insights into phonology (e.g., Clements, 2003) and have also served as the basis for studies on putative extra-linguistic associations with speech (Hay and Bauer, 2007; Atkinson, 2011). Studies of the latter sort have claimed, *inter alia*, that a language's number of phonemes correlates positively with speaker population size and negatively with its distance

E-mail address: caleb@miami.edu.<https://doi.org/10.1016/j.langsci.2018.07.003>

0388-0001/© 2018 Elsevier Ltd. All rights reserved.

Please cite this article in press as: Everett, C., The similar rates of occurrence of consonants across the world's languages: A quantitative analysis of phonetically transcribed word lists, *Language Sciences* (2018), <https://doi.org/10.1016/j.langsci.2018.07.003>

from Africa, though such claims have been argued against compellingly (Moran et al., 2012; Maddieson et al., 2011). Despite such extensive and cross-disciplinary attention paid to phoneme inventories, though, relatively little attention has been paid to the rates at which sound types are utilized in language. This is understandable since large corpora of transcribed texts are available for only a fraction of the world's languages. Some research has investigated the rates of occurrence of consonants within that fraction of languages (Yegelehner and Voegelin, 1957; Gusein-Zade, 1987). Yet no study has documented, on a global scale, the rates of occurrence of sound types within languages, as opposed to their prevalence across phoneme inventories. This study addresses that lacuna in the literature by analyzing a database of phonetically transcribed word lists representing thousands of languages across every major geographic region and linguistic phylum. Patterns in this database suggest that, despite their well-known phonetic and phonological diversity, languages are often quite similar vis-à-vis the rates at which certain phones occur in their word lists.

The data analyzed here offer additional support for previous studies that have presented physiological motivations for the pervasiveness of some sounds in speech (These include Locke, 1983, Ohala, 1983, and Lindblom and Maddieson, 1988, *inter alia*). They do so by offering the advantage of examining phones in such a large sample of languages. Phoneticians and phonologists have long been aware that such large samples are beneficial, given that sufficiently representative samples are critical to investigating universal biases in articulation and perception. As Ohala (1980:182) notes, "Another general issue concerns the problem of how to obtain a truly representative sample of sound patterns from a variety of languages such that the sample is not biased by including too many or too few languages having certain genetic, typological, or geographical linkages." Here that problem is addressed. Nevertheless, the limitations of the database used in this study should also be borne in mind from the outset. While the data used here offer some clear advantages given the number of language varieties represented, some conclusions are tempered because each language is only represented with a small phonetic sample. Still, the data are suggestive of strong tendencies in the usage of consonants, across the world's language families. The particular tendencies focused upon in this study are generally consistent with previous work in the literature that is based on very different methodologies and databases. The distributional tendencies in question are used to quantify the similarity of consonant usage across languages, and to derive a way of isolating languages and linguistic phyla that are typologically anomalous with respect to their usage of consonants. The patterns uncovered here, and the method introduced, merit further exploration with larger corpora for individual languages.

2. Data and general analysis

The database relied on for this study is the Automated Similarity Judgment Program (ASJP), which contains transcriptions of words that are resistant to borrowing in thousands of languages (For the list of 40 words found across all lists, see https://en.wikipedia.org/wiki/Automated_Similarity_Judgment_Program). Most lists in the database represent 40 concepts that are essentially a subset of the 100-word Swadesh list, though some lists have all 100 words (Swadesh, 2006). This database was developed to investigate the relatedness of languages, though it has been used for other purposes also (Wichmann et al., 2016; Brown et al., 2013; Blasi et al., 2016). There are presently over 7000 word lists in the database, though this study focuses on 6901 word lists for distinct dialects and languages. It excludes, for example, word lists based on constructed languages. A map depicting the locations of the represented language varieties is provided in Fig. 1.

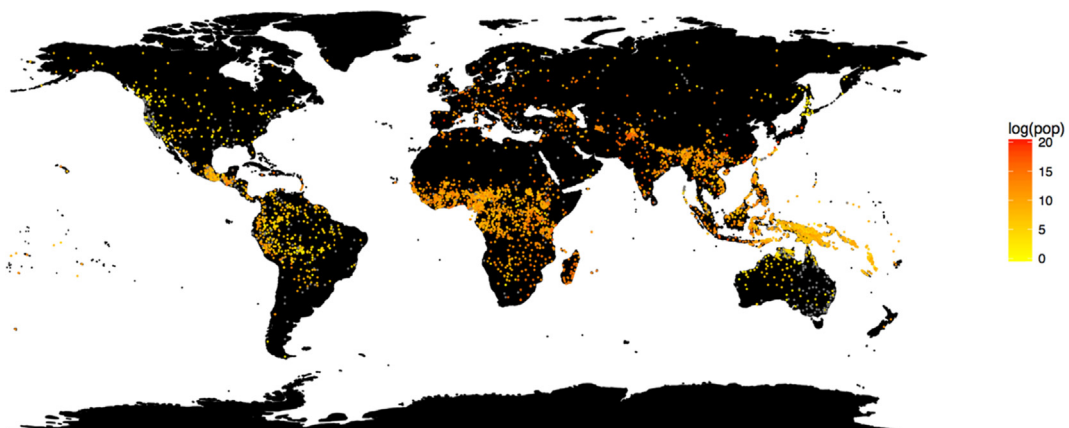


Fig. 1. Locations of the language varieties examined for this study, with log of speaker population. Extinct varieties are represented with gray dots.

There are 4571 distinct ISO codes represented in the data, so the majority of the world's approximately 7000 extant languages are represented. (ISO refers to the international standard of language codes, so the presence of 4571 distinct codes suggests that there are that many mutually unintelligible languages in the data.) Many languages are represented with word lists for more than one dialect. For that reason it is particularly critical to control for relatedness. These languages were grouped into families by relying on one commonly utilized language taxonomy, that of the World Atlas of Language Structures

Download English Version:

<https://daneshyari.com/en/article/8942559>

Download Persian Version:

<https://daneshyari.com/article/8942559>

[Daneshyari.com](https://daneshyari.com)